

ANEXO 3: METODOLOGIA DE ANÁLISIS ESTADÍSTICO DE LA INFORMACIÓN

Para el análisis estadístico de la información recolectada se usó el software estadístico SPSS con el objetivo de analizar las diferentes variables presentes en la encuesta aplicada, la posible dependencia entre variables y relaciones existentes entre estas.

Para este estudio de variables se hizo un análisis individual realizando el cálculo de los estadísticos descriptivos y tablas de frecuencia, posteriormente se realizó un análisis conjunto con el fin de buscar relación entre los diferentes grupos de datos y así poder identificar asociaciones o comportamientos a tener en cuenta en la estructuración del modelo y descripción del sector lácteo.

A continuación se describen los tipos de variables encontrados en la encuesta que pueden ser según el tipo de pregunta, las alternativas de respuesta y el orden jerárquico:

- Tipo de pregunta: Variables categorías o cualitativas, cuyas opciones de respuesta son una características o cualidad que no pueden ser medidas con números.
- Alternativas de respuesta: Dicotómicas (2 alternativas), politómicas (más de 2 alternativas).
- Orden intrínseco: variables nominales que no tienen un orden jerárquico entre ellas (género), y variables ordinales que incluyen un orden jerárquico determinado.

El grado de relación entre variables depende de la naturaleza de las variables implicadas en la investigación, en este sentido si ambas variables son nominales y categóricas la relación será descrita por el estadístico Ji-Cuadrado, si ambas son ordinales se describe con el coeficiente de correlación de Spearman, si ambas variables son intercalares el análisis se realizara a través del coeficiente de Pearson. A continuación se describe el tratamiento estadístico que se realizó a las variables categóricas y a las cualitativas:

ANÁLISIS DE VARIABLES CATEGÓRICAS

Para el análisis de las variables cualitativas conocidas también como variables categóricas es necesario el uso de tablas de contingencia, esta es una tabla de doble entrada con i filas y j columnas, siendo i y j el número de categorías de cada uno de las variables o atributos, las casillas de intersección muestran la frecuencia observada absoluta que presentan simultáneamente las correspondientes filas y columnas. (Calderon and Alzamo 2011)

Para determinar la relación entre dos variables usamos la prueba ji cuadrado χ^2 (Chi cuadrado), que sirve para someter a prueba de hipótesis distribuciones de frecuencias, En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula. Para establecer la existencia de relación entre variables es necesario establecer una prueba de independencia, planteando así una hipótesis nula que postula la independencia entre las variables, lo que indica que no existe diferencia estadísticamente significativa entre las frecuencias observadas y frecuencias esperadas.

Variable A	Variable B				n _{i.}
	B ₁	B ₂	...	B _k	
A ₁	n ₁₁	n ₁₂		n _{1k}	n _{1.}
A ₂	n ₂₁	n ₂₂		n _{2k}	n _{2.}
...		
A _m	n _{m1}	n _{m2}		n _{mk}	n _{m.}
n _{.j}	n _{.1}	n _{.2}	...	n _{.k}	N

Tabla 3.1. Tablas de contingencia de doble entrada, n_{ij} = número de observaciones que tienen el atributo i y j . (Frecuencias observadas), $n_{i.}$ = número de individuos que tienen el atributo i (marginal i), $n_{.j}$ = número de individuos que tienen el atributo j (marginal j)

PRUEBA CHI CUADRADO

Partiendo de la tabla de contingencia inicial, se puede estimar la probabilidad de una variable a partir de su frecuencia (ver ecuación 1), para hallar la probabilidad conjunta de dos variables se parte de la suposición que las dos variables son independientes ($P_i \cdot P_j$) y por tanto E_{ij} será la frecuencia esperada o teórica (ver ecuación 2) y el número esperado de elementos de cada casilla.

$$P_{ij} = \frac{n_{ij}}{N} \quad P_{i.} = \frac{n_{i.}}{N} \quad P_{.j} = \frac{n_{.j}}{N}$$

$$E_{ij} = \frac{n_{i.} \times n_{.j}}{N}$$

Ecuación 3.1

Pearson planteo el estadístico χ^2 (Ver ecuación 3) para analizar la independencia partiendo de una hipótesis nula a contrastar que será la de independencia entre los factores, siendo la hipótesis alterna la de dependencia entre los factores (Otero & Medina Moral 2005) en χ^2 la discrepancia o diferencia entre las frecuencias observadas y las frecuencias esperadas permite concluir que si hay poca diferencia entre valores las variables o atributos serán independientes.

H₀: Las dos variables son independientes. (No existe relación entre las variables)

H₁: Las dos variables no son independientes. (Existe relación entre las variables)

$$\chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

Ecuación 3.2

El valor χ^2 calculado se compara con el valor tabulado en tablas de una χ^2 para un nivel de confianza, en su mayoría 0,05 y los grados de libertad determinados por el número de filas y columnas de las variables o atributos analizados $v = (m - 1)(k - 1)$.

Si el valor calculado es mayor que el valor en tablas de una $\chi^2_{\alpha (m-1)(k-1)}$, significará que las diferencias entre las frecuencias observadas y las frecuencias teóricas o esperadas son muy elevadas y por tanto diremos con un determinado nivel de confianza que existe dependencia entre los factores o atributos analizados. (Otero & Medina Moral 2005)

Resumiendo:

$\chi^2 > \chi^2_{\alpha (m-1)(k-1)}$, Rechazar hipótesis nula (dependencia entre las variables)

$\chi^2 < \chi^2_{\alpha (m-1)(k-1)}$, Aceptar hipótesis nula (independencia entre las variables)

Ecuación 3.3

ANÁLISIS DE VARIABLES CUANTITATIVAS

Para medir la asociación o relación entre variables cuantitativas se usa el coeficiente de correlación de Pearson, este mide la asociación lineal de las variables, es una prueba paramétrica para variables numéricas con distribución normal, el signo del coeficiente está acotado entre -1 y +1, (“COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON,” n.d.) su signo indica la dirección positiva o negativa de la asociación lineal, y su valor absoluto indica la correlación alta entre dos variables, si esta es una asociación lineal perfecta toma el valor de 1, cuando a medida que aumenta una variable aumenta la otra, y perfectamente negativa a medida que aumenta una variable disminuye la otra. (Calderon and Alzamoá 2011)

Se parte de un supuesto o hipótesis nula de independencia entre variables y una hipótesis alternativa de dependencia entre variables, buscando probar la hipótesis alternativa, para ello es necesario hacer el análisis de datos calculando el coeficiente de correlación de Pearson.

H_0 : Hipótesis nula. Las variables en estudio son independientes.

H_1 : Hipótesis Alterna, Las variables en estudio están relacionadas

El coeficiente de correlación de Pearson es un valor adimensional y es la covarianza sobre la desviación típica de las variables analizadas y viene dado por la ecuación x, existen diferentes escalas para interpretar el resultado del coeficiente de correlación de Pearson.(“Análisis de Correlación Simple, Múltiple, Parcial,” n.d.)

$$r_{xy} = \frac{S_{xy}}{S_x S_y} = \frac{\sum_{i=1}^N (x_i - \bar{x}) - (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}}$$

Ecuación 3.4

Magnitud de Correlación	Significado
-1,0	Correlación negativa perfecta
-0,9	Correlación negativa fuerte
-0,75	Correlación negativa considerable
-0,5	Correlación negativa media
-0,1	Correlación negativa débil
0	Correlación nula
0,1	Correlación positiva débil
0,5	Correlación positiva media
0,75	Correlación positiva considerable
0,9	Correlación positiva fuerte
1	Correlación positiva perfecta

Tabla 3.2. Magnitud de correlación.

Una vez se calcula el coeficiente de correlación r es necesario mostrar si hay verdadera relación entre las variables analizadas o tan solo presentan dicha relación como consecuencia del azar, para ello se analiza la significación de dicho coeficiente de correlación, que lo acompaña. Un coeficiente de correlación es significativo si se puede afirmar, con una cierta probabilidad, que es diferente de cero. (“COEFICIENTE DE CORRELACIÓN LINEAL DE PEARSON,” n.d.).

Esta significación viene dada por un valor p , que es el riesgo que se corre al rechazar la hipótesis de independencia con la información que proporcionan los datos, si este valor es 0.05 (riesgo bajo) concluimos que la correlación es significativa y hay una relación real, no debida al azar y por tanto se rechaza la hipótesis nula, si es un valor de 0.01 concluimos que la correlación es altamente significativa (“Análisis de Correlación Simple, Múltiple, Parcial,” n.d.)